



García, María del Carmen

Rapelli, Cecilia

Instituto de investigación Teóricas y Aplicadas de la Escuela de Estadística.

UNA COMPARACIÓN DE MÉTODOS DE DIAGNÓSTICOS PARA LOS MODELOS LINEALES MIXTOS¹

Resumen:

Los modelos lineales mixtos se utilizan para el análisis de datos longitudinales. La elección de un modelo adecuado para los datos depende, necesariamente, del uso de herramientas de diagnóstico para mostrar alguna deficiencia en el modelo estimado. Se justifica su uso ya que en muchos casos hay poca o ninguna base teórica disponible para sugerir la forma específica de las interrelaciones de las variables intervinientes. Estos métodos pueden ayudar a mejorar la elección del modelo y a identificar valores atípicos o unidades influyentes que merecen una mayor atención por parte del investigador. Aunque el modelo mixto ofrece gran flexibilidad para modelar la correlación dentro de la unidad, frecuentemente presente en datos con mediciones repetidas, sufre de la misma falta de robustez frente a observaciones distantes que otros modelos estadísticos basados en la distribución gaussiana. En este trabajo se presentan diferentes estrategias que ayudan a la modelación, introduciendo tanto los enfoques clásicos, que son habitualmente utilizados, como los de reciente aparición. Para la aplicación se utiliza un conjunto de datos, clásico en el análisis de datos longitudinales, que fueron recolectados para evaluar el cambio a través del tiempo de la longitud maxilar de varones y mujeres.

Palabras claves: Datos longitudinales Modelos mixtos Diagnósticos

Abstract:

Mixed linear models are used for longitudinal data analysis. The choice of a suitable model for the data depends, necessarily, on the use of diagnostic tools to show some deficiency in the estimated model. They are used in many cases where there is little or no theoretical basis available to suggest the specific form for the variable's relation. These methods can improve the model selection process, the outlier detection and the detection of influential units that deserve greater attention from the researcher. Although the mixed model offers great flexibility for modeling the correlation within the unit, often present in data with repeated measurements, it suffers from the same lack of robustness to distant observations as other statistical models based on the Gaussian distribution. In this paper, different strategies that help modeling are introduced, using, both, the classical approaches, which are commonly used, and those recently proposed. In the application stage a classic data set used for longitudinal analysis is used to evaluate the change over time of the maxillary length of males and females.

Keywords: Longitudinal data Mixed models Diagnostics

1. Introducción

¹ Este trabajo se elaboró en el marco del Proyecto ECO214 Titulado "Métodos estadísticos avanzados en investigación en ciencias biológicas", dirigido por Cristina Cuesta



La comprobación de la aptitud de un modelo es un paso esencial del modelado estadístico para garantizar los supuestos necesarios para una inferencia válida.

Los análisis de residuos y de sensibilidad constituyen herramientas importantes para evaluar el ajuste de cualquier modelo estadístico. El primero verifica la validez de sus supuestos, y en consecuencia, la confiabilidad de la inferencia estadística basada en él. Por otro lado, el análisis de sensibilidad se basa en un conjunto de herramientas diseñadas para evaluar los cambios en el modelo ajustado cuando se impone alguna perturbación sobre los datos o sobre los supuestos del mismo.

En este trabajo se pretende ilustrar el uso de los análisis mencionados para comprobar el cumplimiento de supuestos o identificar valores y unidades atípicos y/o influyentes.

2. Modelos lineales mixtos

En los estudios longitudinales las unidades se observan repetidamente en varias ocasiones. Los modelos lineales mixtos que contienen efectos fijos y aleatorios se utilizan para el análisis de este tipo de datos.

El modelo lineal mixto se puede expresar como,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon} \quad (2.1)$$

siendo,

$\mathbf{Y} = (\mathbf{Y}'_1, \mathbf{Y}'_2, \dots, \mathbf{Y}'_N)'$, $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_N)'$, $\mathbf{b} = (\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_N)'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_N)'$ y $\mathbf{Z} = \bigoplus_{i=1}^N \mathbf{Z}_i$ donde \bigoplus representa la suma directa, \mathbf{Y}_i , $i=1, \dots, N$, el vector $(n_i \times 1)$ de medidas repetidas del i -ésimo individuo, \mathbf{X}_i y \mathbf{Z}_i son matrices conocidas de dimensión $(n_i \times p)$ y $(n_i \times k)$ respectivamente, \mathbf{b}_i es un vector aleatorio de dimensión $(k \times 1)$, cuyas componentes se denominan efectos aleatorios, $\boldsymbol{\varepsilon}_i$ es un vector aleatorio $(n_i \times 1)$ de mediciones de error (intra-sujeto), $\boldsymbol{\beta}$ un vector $(p \times 1)$ de parámetros.

Los supuestos se escriben,

$$\begin{pmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N_{kN+n} \left(\begin{pmatrix} \mathbf{0}_{kN} \\ \mathbf{0}_n \end{pmatrix}, \begin{pmatrix} \sigma^2 \mathbf{D} & \mathbf{0}_{kN \times n} \\ \mathbf{0}_{n \times kN} & \sigma^2 \boldsymbol{\Sigma} \end{pmatrix} \right),$$

con, $n = \sum_{i=1}^N n_i$, $\mathbf{D} = \mathbf{I}_N \otimes \mathbf{G}$ y $\boldsymbol{\Sigma} = \bigoplus_{i=1}^N \mathbf{R}_i$, \otimes denota el producto Kronecker e \mathbf{I}_N la matriz identidad

de orden N y además, $\mathbf{b}_i \stackrel{iid}{\sim} N_k(\mathbf{0}, \sigma^2 \mathbf{G})$ y $\boldsymbol{\varepsilon}_i \stackrel{ind}{\sim} N_{n_i}(\mathbf{0}, \sigma^2 \mathbf{R}_i)$. Bajo el modelo (2.1) la matriz de variancias y covariancias de \mathbf{Y} , cuyos elementos se sintetizan en un vector $\boldsymbol{\theta}$ de parámetros de covariancia, es $\text{Var}(\mathbf{Y}) = \mathbf{V} = \sigma^2 (\mathbf{Z}\mathbf{D}\mathbf{Z}' + \boldsymbol{\Sigma})$.

Los métodos de máxima verosimilitud (ML) y máxima verosimilitud restringida (RML) pueden ser utilizados para estimar los parámetros y los componentes de varianza del modelo. La estimación de los parámetros implica el uso de métodos iterativos para maximizar la función de verosimilitud, dada su naturaleza no lineal.

Condicional a \mathbf{D} y $\boldsymbol{\Sigma}$ el estimador máximo verosímil de $\boldsymbol{\beta}$ es,



$$\hat{\beta} = \left(\sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1}(\theta) \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{V}}_i^{-1}(\theta) \mathbf{Y}_i,$$

y la predicción del vector de efectos aleatorios, denominada EBLUP,

$$\hat{\mathbf{b}} = \hat{\mathbf{D}} \mathbf{Z}' \hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}).$$

Los métodos de estimación basados en la función de verosimilitud son sensibles a unidades atípicas y su presencia puede tener una gran influencia sobre los resultados del análisis.

3. Métodos de diagnóstico

Los métodos de diagnóstico comprenden técnicas cuyo propósito es investigar la flexibilidad y robustez de los supuestos realizados al elegir un modelo. Es posible dividir las técnicas en dos tipos: análisis de residuos, que investiga los supuestos sobre la distribución de errores y la presencia de valores atípicos y análisis de sensibilidad, que analiza los cambios en el ajuste del modelo cuando está sujeto a perturbaciones menores.

3.1. Análisis de residuos

Un análisis de residuos se debe emplear para chequear la validez de los supuestos subyacentes. Puesto que los modelos mixtos incluyen dos fuentes de variabilidad (intra y entre unidad) se definen tres tipos de residuos (Hilden-Minton, 1995),

- i. **Residuos marginales**, $\hat{\xi} = \mathbf{Y} - \mathbf{X}\hat{\beta} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{Y} = [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}]\xi$, que predicen los errores marginales, $\xi = \mathbf{Y} - \mathbf{E}(\mathbf{Y}) = \mathbf{Y} - \mathbf{X}\beta$.
- ii. **Residuos condicionales**, $\hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\mathbf{b}} = \Sigma\mathbf{Q}\mathbf{Y} = \Sigma\mathbf{Q}\varepsilon + \Sigma\mathbf{Q}\mathbf{Z}\mathbf{b}$, que predicen los errores condicionales, $\varepsilon = \mathbf{Y} - \mathbf{E}(\mathbf{Y} / \mathbf{b}) = \mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\mathbf{b}$.
- iii. **Residuos de efectos aleatorios**, el BLUP, $\mathbf{Z}\hat{\mathbf{b}} = \mathbf{Z}\mathbf{D}\mathbf{Z}'\mathbf{Q}\mathbf{Z}\mathbf{b} + \mathbf{Z}\mathbf{D}\mathbf{Z}'\mathbf{Q}\varepsilon$, que predice los efectos aleatorios, $\mathbf{Z}\mathbf{b} = \mathbf{E}(\mathbf{Y} / \mathbf{b}) - \mathbf{E}(\mathbf{Y})$,

$$\text{donde } \mathbf{T} = (\mathbf{X}'\mathbf{M}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}, \mathbf{Q} = \mathbf{M}(\mathbf{I} - \mathbf{T}), \mathbf{M} = \sigma^2 \mathbf{V}^{-1} = (\mathbf{Z}\mathbf{D}\mathbf{Z}' + \Sigma)^{-1}.$$

Cada tipo de residuo es útil para evaluar algún supuesto del modelo (2.1). En general, los marginales se usan para estudiar la linealidad de la respuesta con respecto a las variables explicativas y para evaluar si la estructura de covariancias intra unidad es adecuada; los condicionales para comprobar la homocedasticidad y normalidad del error condicional y la presencia de unidades atípicas y los EBLUP para la normalidad de los efectos aleatorios y unidades atípicas.

De acuerdo con Hilden-Minton (1995), un residuo es considerado puro para un tipo de error específico si éste depende sólo de las componentes fijas y del error que se supone que predice. Los residuos que dependen de otros tipos de error se denominan "residuos confundidos". Los residuos $\hat{\varepsilon}$ y $\mathbf{Z}\hat{\mathbf{b}}$ están confundidos con \mathbf{b} y ε , respectivamente. Esto implica, por ejemplo, que $\hat{\varepsilon}$ puede no ser adecuado para chequear la normalidad de ε debido a que cuando \mathbf{b} es no normal, $\hat{\varepsilon}$ puede no presentar un comportamiento normal aun cuando ε sea normal. Hilden-Minton (1995) comentaron que la habilidad para chequear la normalidad de ε , usando $\hat{\varepsilon}$, incrementa a medida que se minimice la "fracción de confusión" para el k-ésimo residuo condi-



cional dada por,

$$0 \leq F_k = \frac{\mathbf{U}_k' \Sigma \mathbf{Q} \mathbf{Z} \mathbf{Z}' \mathbf{Q} \Sigma \mathbf{U}_k}{\mathbf{U}_k' \Sigma \mathbf{Q} \Sigma \mathbf{U}_k} = 1 - \frac{\mathbf{U}_k' \Sigma \mathbf{Q} \Sigma \mathbf{Q} \Sigma \mathbf{U}_k}{\mathbf{U}_k' \Sigma \mathbf{Q} \Sigma \mathbf{U}_k} \leq 1 \quad (2.2)$$

donde $\mathbf{U}_k = (\mathbf{u}_{ij})_{n \times v} = (\mathbf{u}_{k_1}, \mathbf{u}_{k_2}, \dots, \mathbf{u}_{k_v})$ siendo \mathbf{u}_k la k -ésima columna de \mathbf{I}_n .

Esta fracción representa la proporción de la variabilidad de $\hat{\epsilon}_k$ atribuida a la confusión con el BLUP. La propuesta de Hilden-Minton es considerar una transformación lineal, $\mathbf{L}'\hat{\epsilon}$, tal que $\mathbf{L}'\hat{\epsilon}$ tenga confusión mínima. Esto se obtiene maximizando (2.2) sujeto a la restricción que $\text{Var}(\mathbf{L}'\hat{\epsilon})$ es mayor que 0.

Los residuos condicionales con confusión mínima se pueden estandarizar y usar para comprobar la normalidad mediante un gráfico probabilístico normal.

Si bien es habitual comprobar la normalidad de los efectos aleatorios utilizando un gráfico probabilístico normal ("QQ-plot") de los EBLUP se aconsejó el uso de un "QQ-plot" ponderado de los mismos. Como las variancias de los efectos aleatorios pueden ser diferentes un gráfico ponderado podría captar mejor esta situación y resulta más sensible a ciertos alejamientos del modelo. Como el gráfico ponderado es difícil de implementar, en los últimos años se recomendó realizar un " χ_k^2 QQ-plot" de la distancia de Mahalanobis entre $\hat{\mathbf{b}}_i$ y la $\mathbf{E}(\mathbf{b}_i) = 0$

$M_i = \hat{\mathbf{b}}_i' (\text{Var}(\hat{\mathbf{b}}_i - \mathbf{b}_i))^{-1} \hat{\mathbf{b}}_i$, la cual tiene distribución χ_k^2 si los efectos aleatorios tienen distribución Gaussiana.

3.2. Análisis de sensibilidad (influencia)

La evaluación cualitativa y cuantitativa de la influencia de unidades sobre el análisis se denomina análisis de sensibilidad o influencia. Este análisis tiene por objetivo determinar las unidades y/o las observaciones que producen excesiva influencia en los parámetros estimados, de manera que permitan desarrollar un modelo más adecuado. El estudio se realiza introduciendo cambios en los datos o en las componentes del modelo y evaluando si se producen cambios importantes en los resultados. Usualmente este análisis se realiza mediante un procedimiento denominado influencia global. Éste considera los cambios en los parámetros poblacionales, sin tener en cuenta los parámetros específicos de la unidad. Una dificultad que surge con este enfoque es determinar cuándo los cambios son suficientemente grandes como para realizar posteriores investigaciones, reformulación del modelo o eliminación de los datos.

Recientemente se propuso un nuevo método que no elimina la unidad, denominado gráfico de la suma de cuadrados de los residuos estudentizados (TRSS) (Mun y Lindstrom, 2013). Este método proporciona una mayor información mediante la utilización de residuos modificados y evalúa eficazmente el efecto de unidades y observaciones discordantes en la estimación de parámetros que incluyen componentes de la variancia.

3.2.1. Influencia global

Una forma de verificar la influencia de un grupo de observaciones es omitirlas y observar los cambios en los estimadores. Si se producen grandes cambios la unidad es influyente.

Para medir el impacto sobre el vector de los parámetros de efectos fijos y covarianza estima-



dos se calcula una estadística a partir de las diferencias entre los estimadores de los parámetros con los datos completos y reducidos. Una estadística de este tipo es la distancia de Cook (D de Cook) que se expresa como,

$$D_i(\beta) = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' \text{Var}(\hat{\beta})^{-1} (\hat{\beta} - \hat{\beta}_{(i)})}{\text{rg}(\mathbf{X})} = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{p},$$

donde, $\hat{\beta}_{(i)}$ denota el vector de estimadores después de eliminar las observaciones en la unidad i y $\text{Var}(\hat{\beta})^{-1}$ la inversa de la matriz de covariancias. Cuanto más grande sea esta estadística mayor es la influencia.

También se puede evaluar la influencia sobre los parámetros de covariancias θ ,

$$D(\theta) = (\hat{\theta} - \hat{\theta}_U)' \hat{\Gamma}^{-1} (\hat{\theta} - \hat{\theta}_U),$$

siendo, Γ la matriz de covariancias asintótica de $\hat{\theta}$.

Tan et al. (2001), entre otros, comentan sobre la eficacia limitada de la distancia de Cook como una medida de influencia en el contexto de LMM. En particular, destacan que al estar orientada a la unidad, esta estadística puede no ser conveniente para detectar unidades influyentes en vista de la posición relativa de las observaciones dentro y entre los sujetos y proponen un enfoque condicional basado en medidas de influencia orientadas a la observación. La medida es similar a la distancia de Cook pero condicional a BLUP. La distancia de Cook condicional se define para la j -ésima observación de la i -ésima unidad como,

$$DC_{i,j} = \frac{1}{\sigma^2((N-1)k+p)} \left[(\hat{\beta} - \hat{\beta}_{i(j)})' \mathbf{X}'\mathbf{X}(\hat{\beta} - \hat{\beta}_{i(j)}) + \sum_{i=1}^N (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)})' \mathbf{Z}'_i \mathbf{Z}_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)}) + 2(\hat{\beta} - \hat{\beta}_{i(j)})' \sum_{i=1}^N \mathbf{X}'_i \mathbf{Z}_i (\hat{\mathbf{b}}_i - \hat{\mathbf{b}}_{i(j)}) \right],$$

donde, $\hat{\beta}_{i(j)}$ y $\hat{\mathbf{b}}_{i(j)}$ denotan, respectivamente, el estimador de β y la predicción del efecto aleatorio sin la j -ésima observación de la unidad i .

La distancia condicional se puede escribir, entonces, como la suma de tres componentes,

$$DC_{i,j} = DC1_{i,j} + DC2_{i,j} + DC3_{i,j},$$

La estadística $DC_{i,j}$ mide la influencia de la observación j de la unidad i , mientras que los distintos términos de la ecuación proporcionan, respectivamente, una medida de influencia sobre los efectos fijos, cambios en los efectos específicos de la unidad y una medida de covariación entre un cambio en el perfil promedio y un cambio en el perfil individual relativo al perfil promedio. El valor de $DC3_{i,j}$ debería estar cerca de cero si el modelo es válido.

3.2.2 Gráficos de las sumas de cuadrados de los residuos estudentizados

Los métodos anteriores se focalizan sólo en los cambios de los coeficientes, no teniendo en cuenta la trayectoria de los casos atípicos.

Esta nueva herramienta diagnóstica (Mun y Lindstrom, 2013) introduce una leve modificación a la expresión de los residuos (residuos modificados) y construye la suma de cuadrados de estos residuos. Considerando el concepto que una unidad influyente está alejada de su media, se



define un valor para medir la desviación entre la media específica de una unidad y la media poblacional y otro que considera la distancia entre una trayectoria individual y su media específica. La primera se denomina desviación tipo L (posición) y la segunda tipo S (forma). Estos dos tipos de desviaciones que se pueden examinar simultáneamente.

Un residuo, definido como la diferencia entre la respuesta y la respuesta media estimada, se descompone como la suma de $e_{i,1}$ y $e_{i,0}$, siendo,

$$e_{i,1} = (\text{respuesta} - \text{respuesta media de la unidad}) = Y_i - (X_i \hat{\beta} + Z_i \hat{b}_i)$$

$$e_{i,0} = (\text{respuesta media de la unidad} - \text{respuesta media}) = X_i \hat{\beta} + Z_i \hat{b}_i - X_i \hat{\beta}$$

Estos residuos se usan para calcular las sumas de cuadrados de los residuos, denominadas ,

$RSS_{i,0} = e_{i,0}' e_{i,0}$ y $RSS_{i,1} = e_{i,1}' e_{i,1}$ que contienen información sobre desviaciones tipo L y S, respectivamente.

Las sumas de cuadrados dependen de las unidades de medida y el número de mediciones por unidad por lo que resulta conveniente estandarizarlas,

$$TRSS_{i,0}^* = \frac{RSS_{i,0} - E(RSS_{i,0})}{\sqrt{Var(RSS_{i,0})}} \quad y \quad TRSS_{i,1}^* = \frac{RSS_{i,1} - E(RSS_{i,1})}{\sqrt{Var(RSS_{i,1})}}.$$

Las desviaciones positivas son más preocupantes que las negativas. Las sumas de cuadrados más chicas que su esperanza indican simplemente que el modelo ajusta mejor para esa unidad que para otras. Por lo cual sólo grandes valores positivos son de interés, definiendo $TRSS_{i,0} = \max\{0, TRSS_{i,0}^*\}$ y $TRSS_{i,1} = \max\{0, TRSS_{i,1}^*\}$.

El gráfico TRSS es un diagrama de dispersión entre $TRSS_{i,0}$ y $TRSS_{i,1}$ que muestra unidades discordantes y sus tipos de desviación:

- Cuando los valores de $TRSS_{i,0}$ son grandes y los de $TRSS_{i,1}$ chicos indica que el sujeto i está lejos de la media marginal (desviación tipo L)
- Cuando los valores de $TRSS_{i,1}$ son grandes y los de $TRSS_{i,0}$ chicos sugieren que la unidad puede tener diferentes trayectorias que las otras (desviación tipo S) o diferente estructura de correlación que las otras.

El gráfico TRSS muestra ambos tipos de desviaciones simultáneamente y permiten investigar una unidad sin eliminarla, pues la medida TRSS es una medida de no omisión de casos. Si existen varias unidades discordantes también permite detectarlas visualmente.

Las unidades discordantes y sus tipos de desviación se determinan por la distancia al origen y la dirección en el gráfico TRSS. Es útil tener líneas de referencia para evaluar si una unidad se puede considerar o no atípica. Existen diferentes formas de definirlas a partir de una normal bivariada truncada, con dos niveles de cobertura (95 y 99%). Una de las más utilizadas es la densidad de probabilidad más alta (HPD), que encuentra un valor de corte y un elipsoide en el primer cuadrante para encontrar la probabilidad de cobertura nominal,

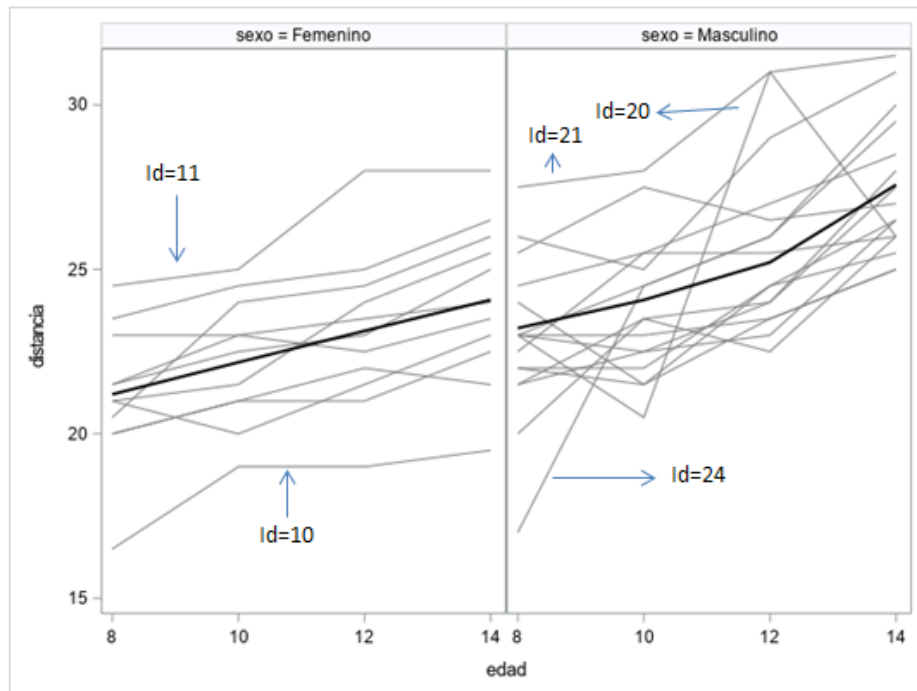
4. Resultados

La metodología descrita se aplica a un conjunto de datos, clásico en el análisis de datos longi-



tudinales, que fueron recolectados en la Universidad de Carolina del Norte y analizados por Pothoff y Roy en 1964. El estudio considera 27 niños, 16 varones y 11 mujeres. Se midió la distancia (en mm) desde el centro de la glándula pituitaria hasta la fisura pterygomaxilar de cada uno de los niños involucrados en el estudio, a las edades de 8, 10, 12 y 14 años, con el objetivo de determinar si la tasa de cambio de la distancia de los maxilares a través del tiempo es similar para varones y mujeres.

Gráfico 1. Evolución de las distancias individual y promedio a través del tiempo por sexo



La inspección de los perfiles individuales de la distancia maxilar en función de la edad muestra que algunas trayectorias están alejadas del resto. El gráfico del perfil promedio sugiere que la trayectoria media es lineal para ambos sexos, con una tasa levemente mayor para los varones.

Se propuso el siguiente modelo lineal mixto con un efecto aleatorio,

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \delta_i + \beta_3 \delta_i t_{ij} + b_{0i} + e_{ij} \quad i = 1, 2, \dots, 27 \quad j = 1, 2, \dots, 4$$

$$\text{Var}(\mathbf{b}_i) = \text{Var}(b_{0i}) = \mathbf{D}$$

$$\text{Var}(\mathbf{e}_i) = \text{Var} \begin{pmatrix} e_{i1} \\ \vdots \\ e_{in_i} \end{pmatrix} = \mathbf{R} = \sigma^2 \mathbf{I}$$

donde, t_{ij} representa la edad y $\delta_i = 0$ si la unidad i es mujer y $\delta_i = 1$ si la unidad es varón i .



La estimación de los parámetros del modelo propuesto se realiza utilizando el paquete “nlme” del software estadístico R.

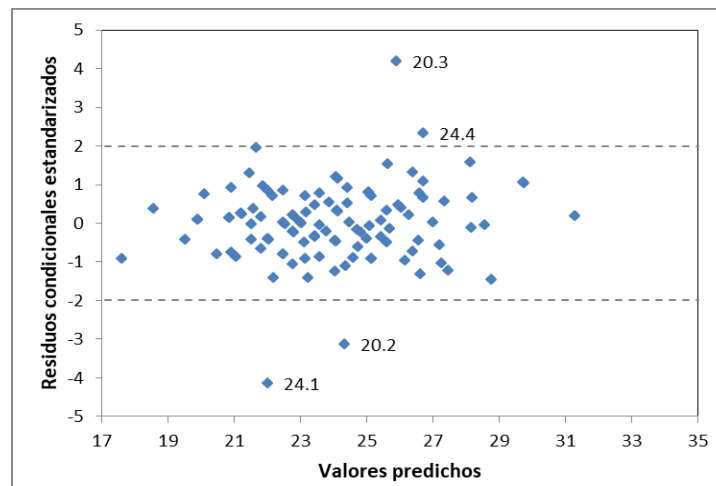
Tabla 1. Estimación de los efectos fijos y sus errores estándares

Parámetro	Estimación	Error Estándar
Ordenada	22.647	0.5861
Edad	0.4795	0.0935
Sexo	2.3210	0.7614
Edad *Sexo	0.3048	0.1214

$$D = \text{Var}(b_{0i}) = 3.298 \quad \sigma^2 = 1.921$$

En primer lugar se realizó un análisis de residuos para el modelo ajustado.

Gráfico 2. Gráfico de residuos condicionales estandarizados vs valores predichos



El gráfico 2 muestra que los residuos condicionales estandarizados se mantienen dentro de un rango sugiriendo que el supuesto de homocedasticidad del error condicional se cumple, aunque se observa la presencia de observaciones atípicas.

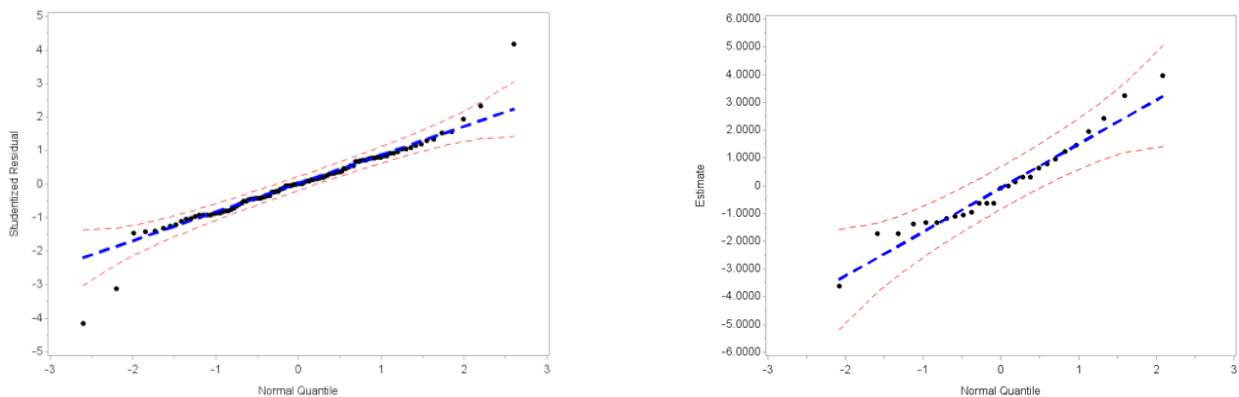
Para evaluar la normalidad del error condicional y de los efectos aleatorios se utiliza un gráfico probabilístico normal (QQ-plot). Si los residuos realmente tienen una distribución normal se espera que el gráfico muestre una línea recta. Se puede graficar una banda de confianza simulada, también llamada “sobre de confianza” (“envelope”), del 95% para el QQ-plot, de manera que si los puntos caen fuera de la banda el supuesto de normalidad para los residuos no es



válido.

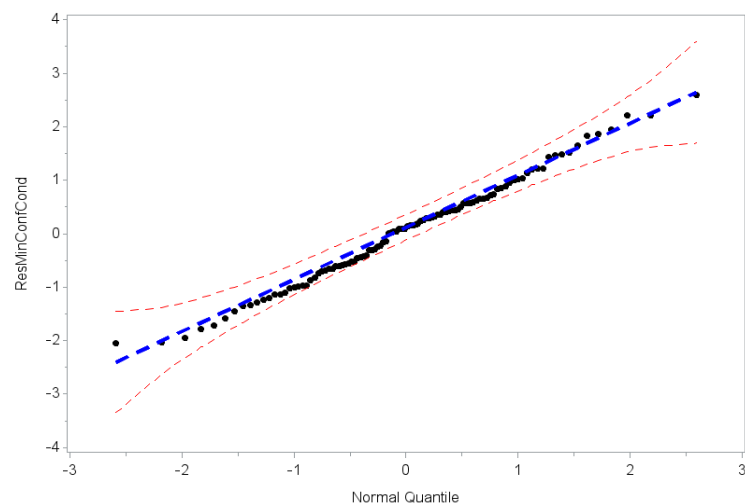
Gráfico 3. Gráfico probabilístico normal con bandas simuladas

(a) Residuos condicionales estandarizados (b) Predicción de efectos aleatorios



El gráfico 3 (a) sugiere no normalidad de los errores condicionales, mientras que en el (b) no se identifican observaciones fuera de la banda indicando normalidad de EBLUP. Sin embargo, el gráfico 3 (a) resulta cuestionable debido al problema de confusión que puede presentarse. Por ese motivo se realiza un QQ-plot de los residuos con confusión mínima.

Gráfico 4. Gráfico probabilístico normal de residuos condicionales con confusión mínima



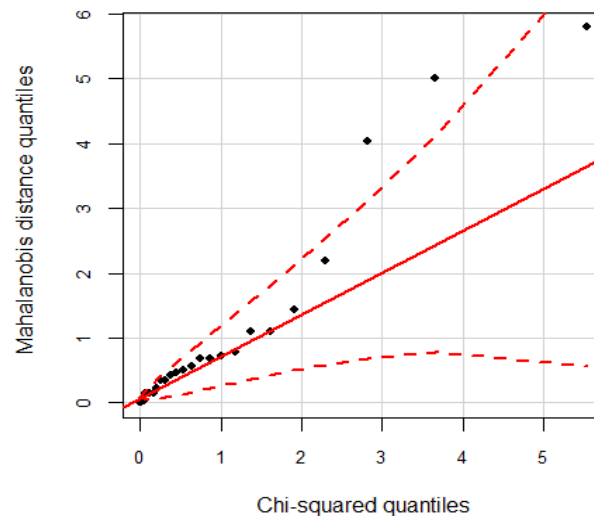
En el gráfico de los residuos condicionales con confusión mínima, no se identifican observaciones fuera del intervalo simulado, indicando que los residuos transformados muestran normalidad.

Como se mencionó anteriormente un gráfico de cuantiles normales ponderado (Lange y Ryan,



1989) para comprobar la normalidad de los efectos aleatorios es dificultoso de implementar (Nobre y Singer, 2007). Para corroborar lo visualizado en el gráfico 3 (b) y considerando la sugerencia de Lange et. al. de que no es conveniente efectuar un gráfico sin ponderar se realiza un " χ^2_k QQ-plot" de la distancia de Mahalanobis para la predicción de los efectos aleatorios.

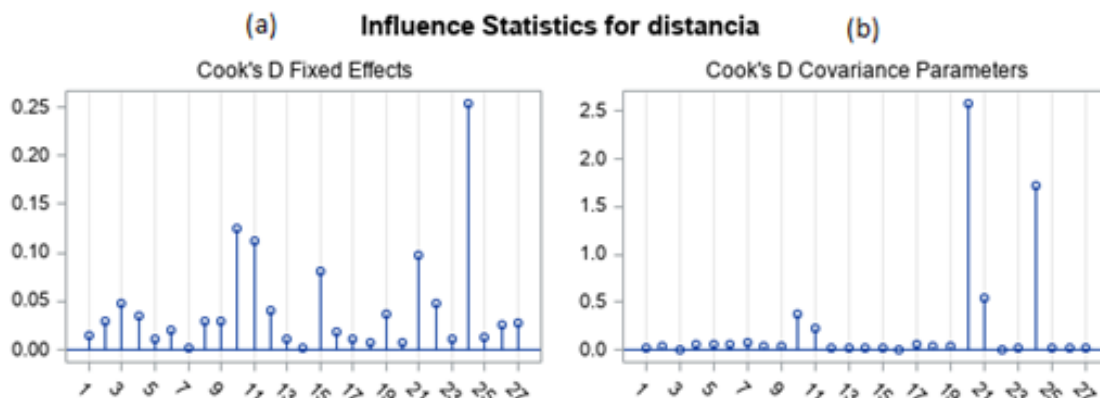
Gráfico 5. Gráfico probabilístico de la distancia de Mahalanobis para los efectos aleatorios



Los puntos que caen fuera de la banda ponen en duda el cumplimiento del supuesto de normalidad de los efectos aleatorios, concluyendo que el gráfico 3 (b) erróneamente muestra normalidad de EA y no normalidad de los residuos condicionales debido a la confusión entre ellos.

Por último se realiza un análisis de sensibilidad.

Gráfico 6. Distancia de Cook para los efectos fijos y parámetros covariancia

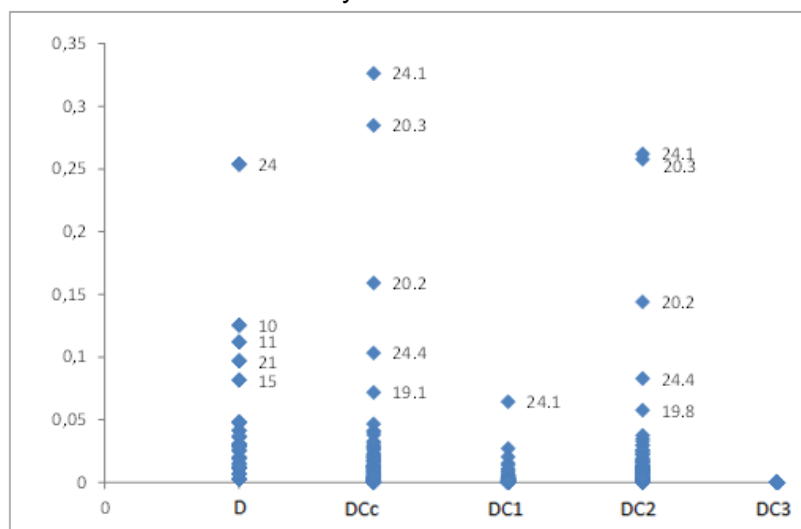




El gráfico de la distancia de Cook, gráfico 6, muestra que el niño con mayor efecto sobre los efectos fijos es el 24. Las niñas 10, 11 y los varones 21 y 15 tienen una influencia fuerte sobre los efectos fijos. La influencia de los individuos 20 y 24 sobre los parámetros de covariancia es notoria, mientras que la del 21 es leve.

Los resultados obtenidos mediante el uso de la distancia Cook condicional y la convencional se resumen en el siguiente gráfico

Gráfico 7. Gráfico de la distancia de Cook y la distancia de Cook condicional



En el gráfico de perfiles individuales se detecta que las unidades 10, 11, 20, 21 y 24 tienen perfiles atípicos. Las cuatro observaciones de la unidad 20 son bastante diferentes a las de las otras unidades, pero su trayectoria específica parece similar a la poblacional para los varones y la distancia de Cook clásica (D) no la identifica pues no causa problemas en la estimación efectos fijos. Sin embargo, esta unidad es detectada en el gráfico 6 (b) (D parámetros covariancia) pues afecta las componentes variancia.

La distancia de Cook detecta a 24 como el más influyente, en lugar de 20, pues su trayectoria es diferente a las demás. Las trayectorias de las longitudes correspondientes a los niños 10, 11 y 21 son similares a la del promedio poblacional pero están alejadas de la misma.

La distancia de Cook condicional mejora el desempeño de D, ya que muestra observaciones (dentro de las unidades) discordantes. Este gráfico muestra como atípicas las distancias máximas del niño 20 a los 8 y 12 años, del niño 24 a los 8 y 14 años y del niño 19 a los 8 años. Esta discordancia se debe mayormente a cambios sobre los efectos aleatorios (DC2).

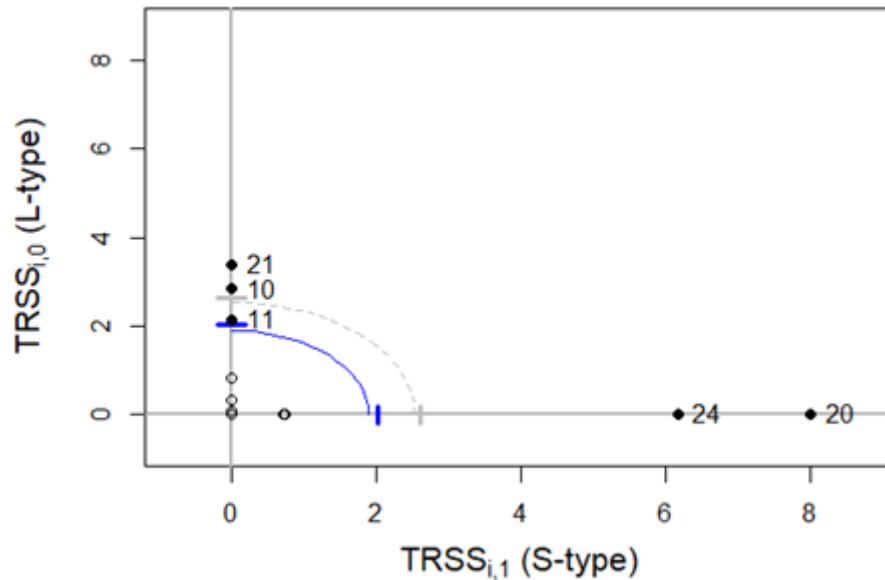
Las unidades 10, 11 y 21 no son detectadas por CD pues su discordancia no se debe a observaciones sino al perfil completo.

Se detecta además que la primera determinación de la longitud maxilar del niño 19 resulta influyente sobre las componentes de variancia.

El gráfico TRSS siguiente muestra las unidades discordantes y las líneas de referencia



Grafico 8. Gráfico de las sumas de cuadrados de los residuos estudentizados



Se observa que las unidades 24, 20, 21 y 10 superan las líneas de referencia del 99% (línea punteada) y 95% (línea sólida), mientras que la unidad 11 es detectada como influyentes mediante la línea del 95%.

Para intentar explicar los motivos de las discrepancias entre los métodos se presentan los gráficos de los perfiles individual y promedio por grupo y los valores observados para las unidades 10, 11, 21 y 20.

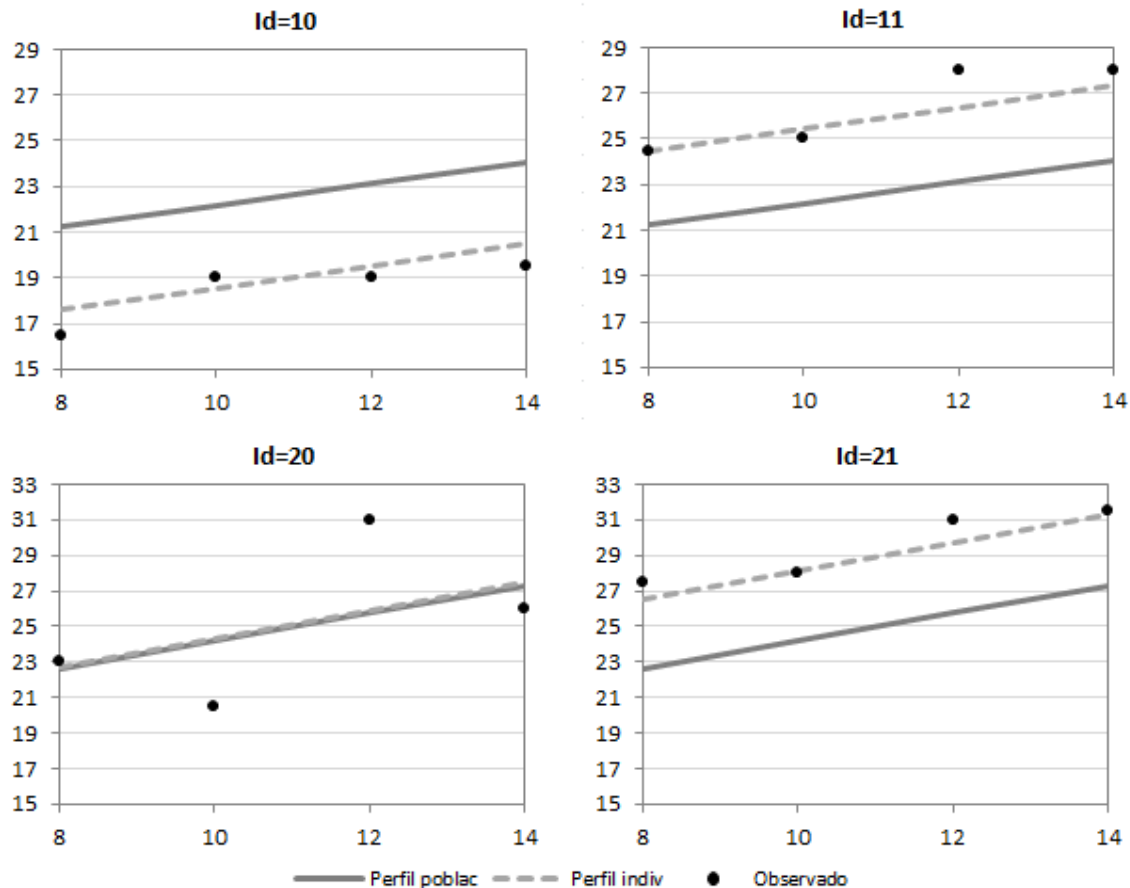
El perfil individual de la unidad 20 es similar al perfil promedio del grupo. Como los valores de $TRSS_{i,1}$ son grandes y los de $TRSS_{i,0}$ chicos, no tiene efectos sobre la estimación de los efectos fijos, por lo cual no la detecta la distancia Cook clásica.

La unidad 21 tiene $TRSS_{i,0}$ grande y $TRSS_{i,1}$ chica sugiriendo que la unidad está lejos de la media marginal, pero debido a que las desviaciones de las observaciones del perfil individual son pequeñas no son señaladas por la estadística Cook condicional.

Las unidades 10 y 11 tienen valores de $TRSS_{i,0}$ grandes y de $TRSS_{i,1}$ chicos indica que el sujeto i está lejos de la media marginal y tiene efectos sobre la estimación de los efectos fijos.



Gráfico 9. Perfiles individuales y promedio por grupo para cuatro pacientes influyentes



4. Discusión

En este trabajo se presentan un análisis de residuos para detectar alejamientos de los supuestos en un modelo lineal mixto y varios enfoques para detectar unidades que tienen una magnitud distinta al resto y el efecto que producen sobre los estimadores de los parámetros del modelo.

Se pueden definir varios tipos de residuos comprobando con cada uno algún supuesto del modelo. Sin embargo, algunos de ellos están sujetos a confusión debido a la presencia de otras fuentes de variación. Específicamente, para evaluar la normalidad se utilizan los residuos condicionales estandarizados que presentan confusión en el sentido que no sólo son función del error que predicen y pueden no ser adecuados para evaluar ese supuesto. Para minimizar el efecto de la confusión se sugiere utilizar una transformación lineal de los residuos condicionados, los residuos con confusión mínima. Para comprobar la normalidad de los efectos aleatorios se utiliza un " χ^2_k QQ-plot" en lugar de un gráfico probabilístico normal convencional.

La idea general del método de influencia global es eliminar observaciones o unidades y evaluar si se producen cambios importantes en los resultados, determinando si se afecta el proceso de



estimación. Los gráficos TRSS no eliminan las unidades ni alteran el modelo para identificar las unidades discordantes. El método proporciona una mayor información sobre las mediciones repetidas mediante la utilización de residuos modificados y evalúa eficazmente el efecto de unidades y observaciones discordantes en la estimación de parámetros que incluyen componentes de la variancia.

En la aplicación se muestra que:

- Los residuos condicionales estandarizados no muestran normalidad.
- Los efectos aleatorios no muestran normalidad.
- Los residuos con confusión mínima evidencian su utilidad para comprobar la normalidad del error condicional.

Esto sugiere que los residuos estandarizados con mínima confusión se pueden emplear para evaluar el supuesto de normalidad del error condicional aun cuando los efectos aleatorios no sean normales.

Con respecto al análisis de influencia, la aplicación sugiere que los métodos presentados se complementan:

- Influencia global: los diagnósticos ayudan a detectar individuos que tienen gran efecto sobre los efectos fijos y otras, casi sin impacto sobre los efectos fijos, se muestran principalmente influyentes sobre los estimadores puntuales de covariancia. Se observa un mejor desempeño de la distancia de Cook condicional que la clásica,
- Gráficos TRSS: detectan en general las mismas unidades que los métodos anteriores. Sin embargo, ayudan a identificar unidades con trayectorias o posiblemente con estructuras de correlación diferentes al resto.

5. Referencias bibliográficas

- Demidenko, E. (2013). *"Mixed models: theory and applications with R"*. Wiley (New York).
- Cook RD. (1977) Detection of influential observation in linear regression. *Technometrics*; 19,15–18.
- Cook, R. D. and Weisberg, S. (1982). *"Residuals and influence regression"*. Chapman & Hall (New York).
- Garcia, M. del C., Koegel, L., Rapelli, C. (2008) Diagnósticos para los modelos lineales mixtos. Un análisis comparativo de dos enfoques para evaluar la influencia. Libro "II Jornada de Ciencia y Tecnología. Divulgación de la Producción Científica y Tecnológica de la UNR". 169-173.
- Garcia, M. del C., Méndez, F. (2007) Métodos diagnósticos para evaluar la influencia en el contexto de los modelos lineales mixtos. XXXV Coloquio Argentino de Estadística. Pág web <http://www.s-a-e.org.ar/ultimos13coloquios.htm>
- Hilden-Minton, J. A. (1995). Multilevel diagnostics for mixed and hierarchical linear models. PhD Thesis. University of California, Los Angeles.
- Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* 54, 570–582.
- Mun, J. and Lindstrom, M. (2013) Diagnostics for repeated measurements in linear mixed ef-



- fects models. *Statistics in Medicine*, 32 1361–1375
- Nobre, J. S. and Singer, J. M. (2006) Residual Analysis for Linear Mixed Models. *Biometrical Journal* 49 (2007) 6, 863-875.
- Pinheiro, J. C. and Bates, D. M. (2000). *"Mixed-effects in S and S-PLUS"*. Springer (New York).
- R Development Core Team (2006). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- SAS Institute Inc. 2004. SAS/STAT Software: Versión 9.1. Cary, NC: SAS Institute Inc.
- Searle, S. R. (1982). *"Matrix Algebra Useful for Statistics"*. Wiley (New York).
- Singer, J. M. , Nobre, J. S. and Rocha, F.M.M. (2016) Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. *International Statistical Review* 49, 1-35.
- Tan F.E.S., Ouwens M.J.N.M. and Berger MPF. (2001) Detection of influential observations in longitudinal mixed effects regression models. *The Statistician*, 50:271–284.
- Waternaux, C., Laird, N. M., and Ware, J. H. (1989). Methods for analysis of longitudinal data: blood-lead concentrations and cognitive development. *Journal of the American Statistical Association* 84, 33–41.
- Weiss, R. E. and Lazaro, C. G. (1992). Residual plots for repeated measures. *Statistics in Medicine* 11, 115–124.